KFPA Monocular Ranging Algorithm Design and Application in Mobile edge Computing

Shuo Chen[1], Songzhu Mei[2], Gangyong Jia[1], Youhuizi Li[1], Weihua Zhao[1]

[1] Department of Computer Science, Hangzhou Dianzi University, China
[2] Science and Technology on Parallel and Distributed Processing Laboratory (PDL), National University of Defense Technology, China
cloud_cs@foxmail.com, sz.mei@nudt.edu.cn, gangyong@hdu.edu.cn, huizi@hdu.edu.cn, computer323@hdu.edu.cn

Abstract

Distance perception is the basis and necessary prerequisite of environment perception, attitude perception and obstacle avoidance for both intelligent vehicle and unmanned vehicle. Passive ranging is a critical part of machine vision measurement. Most passive ranging methods based on machine vision apply binocular technology that needs strict hardware conditions and lacks universality. Therefore, the monocular vision ranging method is one of the mainstream distance sensing methods at present. In order to improve the accuracy of monocular vision ranging, a monocular vision ranging method based on pixel area and aspect ratio is proposed. Subsequently, this method improves the stability of real-time target detection by introducing Kalman filter processing. Experimental results display that that ranging used by this method has higher accuracy. The mean relative error of the depth measurement is 5% when it is 3-10 m. After introducing Kalman filter, the stability of real-time ranging processing is improved by 25.21%. In this paper, the ROS smart car with real-time target tracking is also realized by the method based on the combination of SIFT-KCF target detection and tracking and monocular ranging.

Keywords: Monocular ranging, Kalman filter, SIFT, KCF, Mobile edge computing

1 Introduction

From December 2019, the outbreak of the "COVID-19" greatly affected the daily lives of people in various countries. In order to effectively control the spread of the epidemic, Chinese government requires residents to stay at home, so as to reduce the flow of people, requires companies to push their late return to work date, and requires schools to postpone the start date of school. The government's effective measures have largely controlled the development of the epidemic. In order to reduce personnel contact during the epidemic prevention period, a number of remote/automatically controlled mobile devices appeared, such as robots delivering meals and drones for shouting [1]. These devices not only facilitate people's lives, but also play a certain role in terms of controlling the epidemic. In order to realize the above applications, the development and application of image processing, mobile network [2] and mobile edge computing [3-6] technology are very important.

In order to ensure stability and effectively control and avoid equipment damage, the above-mentioned equipment needs to be equipped with corresponding ranging equipment. The data obtained by the machine through the distance sensor is converted into corresponding control commands, and then a series of control commands based on distance are completed. There are many options in ranging methods:

Ultrasonic distance measurement [7] includes the earliest application and the most mature technology. Its working principle is relatively simple and has a good ranging effect within a certain distance. Ultrasonic distance measurement is based on the principle that ultrasonic waves could be reflected when they encounter obstacles during propagation.

Microwave radar [8] has the characteristics of small size, long measuring distance, high precision and high reliability. Microwave radar can detect multiple targets at the same time, and the radar range is less affected by climatic conditions. The target can be found in time under severe conditions, so as to ensure control safety.

Laser ranging [9] is relatively simple in principle. Its working principle is similar to that of microwave radar ranging. There are two ranging methods, namely continuous wave and pulse wave. The advantages of laser ranging are short measurement time, high range and good accuracy.

Infrared ranging [10] applies infrared to carry out ranging. It is mainly suitable for dark night or low visibility environment. Infrared ranging has various

*Corresponding Author: Songzhu Mei; E-mail: sz.mei@nudt.edu.cn
DOI: 10.53106/160792642021092205016

advantages, including small size, light weight, low power consumption and safety.

Visual ranging has a wide range of application scenarios [11-13]. The computer processes, analyzes and calculates the video or images that are collected by the camera monitoring. Then the computer uses the established mathematical model to calculate the distance. Compared with several algorithms introduced above, the biggest feature of visual ranging refers to passive ranging that does not need to transmit any signal or receive signal to the measured object. Visual ranging only needs to shoot a video or picture that contains the desired target, so as to measure the distance of the target as required.

According to the number of visual imaging equipment used, visual ranging methods [14] can be divided into monocular vision ranging methods, binocular vision ranging methods (stereoscopic vision) [15] and multi-eye vision ranging methods (omnidirectional vision) [16]. Due to the actual factors such as installation platform, venue and price, the actual application ratio of monocular vision is far greater than that of binocular and multi-eye ranging. The monocular vision ranging method is to apply only one visual imaging device to collect images, and to measure the geometric size of the target, the position and posture of the target in space, and other information. This paper mainly studies how to solve the problem of multi-angle real-time ranging based on specific targets and under the condition of monocular vision in the field of computer vision.

In order to realize the intelligent target tracking smart car system, aiming at the problems in the tracking process, this thesis proposes a multi-angle adaptive KFPA monocular ranging algorithm. The main contributions of this article are as follows:

(1) Proposing a target detection and tracking algorithm based on SIFT-KCF to achieve effective target detection and tracking.

(2) Due to the different target images at different angles, this paper proposes a monocular ranging method on the basis of the target aspect ratio.

(3) Different from relying solely on width or height, the pixel area is introduced as the basis of judgment, so as to improve the overall anti-interference stability.

(4) On the basis of the above, Kalman filter is adopted to improve the stability of real-time moving target ranging.

The rest of this article is structured as follows. The related work in this field is introduced in Section 2, and the real-time ranging processing method is introduced in Section 3. In Section 4, the parameters and architecture of the experimental platform are introduced, while Section 5 analyzes the experimental data and outlines our future work.

2 Related Work

In recent years, with the improvement of hardware performance and the diversity of people's needs [17-18], Ranging technology is widely used in applications such as autonomous driving and drones. This article introduces ranging related work from two aspects: Ranging method based on multiple sensors and Ranging method based on monocular vision.

2.1 Ranging Method Based on Multiple Sensors

Recently, smart cars and drones have a large number of sensors, so the method of multi-sensor data fusion is widely applied. In order to improve the measurement accuracy of the binocular vision sensor in the target tracking process, Wang et al. [19] proposed a method for auxiliary correction with a laser ranging sensor. In order to solve the problem that the traditional pedestrian recognition and ranging effect is not ideal, the literature puts forward a monocular vision night pedestrian recognition and fast ranging method that is suitable for infrared auxiliary night vision images. The fusion of inertial measurement unit (IMU) sensors [20], ultrasonic sensors [21], microwave sensors [22] and cameras are used in devices, including smart cars and drones.

2.2 Ranging Method Based on Monocular Vision

Compared with the additional sensor method, the monocular ranging is simple, the amount of calculation is small, the cost is low and the real-time performance is better.

In recent years, many achievements have been made in the research of monocular vision ranging methods at home and abroad [23-26], and they are mainly divided into fitting modeling method, inverse perspective transformation method, imaging geometric relationship method and optical projection characteristic method. Xu et al. [27] came up with a feature transformation matching algorithm to estimate the attitude angle of the tank, thus using the matched template and target image to simulate stereo vision. This method transforms the monocular distance measurement into a binocular vision plane distance measurement method, but requires a large number of attitude angle matching templates. Poor template matching could cause large ranging errors and consume a lot of matching time.

Han et al. [28] proposed a monocular vision vehicle ranging method based on vehicle width and length estimation, and comprehensively considered two ranging environments, structured roads and unstructured roads. The literature adopts Kalman filter algorithm to reduce the ranging error caused by the change of the vehicle pitch angle. However, this method has a large error in the estimation of the

vehicle width and length of the non-front target, and it does not model and analyze the mechanism of the pitch angle change. It only reduces the estimation error of the vehicle width and length in the tracking process. Bao et al. [29] proposed a monocular vision vehicle ranging method based on the linear relationship between the average length of the vehicle width and the actual distance of the detected vehicle. However, it does not take the attitude angle changes during the movement of the car into account, and the average vehicle width obtained by statistics can only guarantee the average ranging accuracy. It has a large ranging error for a single vehicle target.

3 Methodology

The method mentioned in this paper is mainly composed of three parts: real-time target detection based on SIFT and KCF, ranging method based on target aspect ratio and real-time measurement based on Kalman filter.

3.1 Real-time Target Detection Based on SIFT and KCF

This paper proposes a target detection and tracking processing method that combines SIFT and KCF to provide effective support for subsequent ranging. The algorithm 1 is as follows:

Firstly, the feature points of the input video sequence and target image are extracted by SIFT algorithm. Secondly, the target feature points and video image feature points are matched by klann to realize the target framing. Finally, the KCF algorithm is utilized to track the target. In the process of tracking, the algorithm detects the effectiveness of tracking. If the APCE [30] is greater than the threshold value tf , it means that the tracking is accurate. Continuing to run the KCF tracking algorithm. If APCE is less than the threshold tf , it marks that the target is lost. Restart SIFT algorithm to frame the target.

$$APCE = \frac{\left|F_{\max} - F_{\min}\right|}{mean(\sum_{w,h}(F_{w,h} - F_{\min})^2)} \quad (1)$$

In the formula: $F_{\max}$, $F_{\min}$, $F_{w,h}$ respectively represent the maximum, minimum and (w,h) response values in the response graph.

Algorithm 1. SIFT and KCF target detection and tracking algorithms.

Require:

Target image: $image_{target}$

Real time video image: $video_{frame}$

Ensure:

Real time target location: $pos = (x, y, W, H)$

1. if $frame == 1$ then
2. $target_{kp} = sift(image_{target})$
3. $video_{kp} = sift(video_{frame})$
4. $pos_{frame} = FLANN(target_{kp}, video_{kp})$
5. $KCF_{start}(pos_{frame})$
6. else
7. while $video_{frame} > 1$ do
8. if $pos_{frame} == null$ then
9. $target_{kp} = sift(image_{target})$
10. $video_{kp} = sift(video_{frame})$
11. $pos_{frame} = FLANN(target_{kp}, video_{kp})$
12. $KCF_{start}(pos_{frame})$
13. else
14. $APCE = \frac{\left|F_{\max} - F_{\min}\right|}{mean(\sum_{w,h}(F_{w,h} - F_{\min})^2)}$
15. if $APCE > tf$ then
16. $pos_{frame+1} = KCF_{next}(pos_{frame})$
17. else
18. $pos = null$
19. end if
20. end if
21. end while
22. end if

The APCE value indicates the relationship between the peak value of the response graph and the response value of each point. A large number of experiments showed that the KCF tracking response graph fluctuates when the tracking is blocked, blurred, or lost in a complex environment. The fluctuation analysis of the response graph can measure the target environment and tracking performance, indicating that the calculation of the APCE value of the current frame can reflect the confidence of the tracking result to a certain extent, and determine the tracking situation of the target.

3.2 The Kalman Filter Algorithm

Kalman filtering is a linear filtering method that could minimize the mean square error. It uses an iterative method to predict and update the target state based on the observation equation and the target state transition equation.

After detecting and marking the target, this paper introduces the Kalman filter prediction algorithm. According to the target detection rectangle aspect ratio and the distance detection model of the pixel area, the prediction information of the target in the next frame is established. Therefore, when the target angle changes and the distance changes, continuous and stable real-time ranging can be established based on the prediction information. The state transition equation and the

observation equation are the basis of Kalman filtering. The formulas are as follows:

$$x_k = A * x_{k-1} + w_{k-1} \tag{2}$$

$$z_k = H * x_k + v_k \tag{3}$$

Among them, z_k represents the observation state of the system at time k, and x_k and x_{k-1} correspond to the state of the system at time k and $k-1$. A refers to the state transition matrix, corresponding to the movement mode of the target. H is the observation matrix, corresponding to the relationship between the internal state and the observation value. v_k corresponds to the measurement noise, obeys the normal distribution, and its covariance is R. w corresponds to the system noise, which also satisfies the normal distribution, and its covariance is Q. The prediction and update equations are as follows:

$$\hat{x}_{k|k-1} = A_{k|k-1}\hat{x}_{k-1} \tag{4}$$

$$P_{k|k-1} = A_{k|k-1}P_{k-1}A_{k|k-1}^T A_{k|k-1}^T + Q_{k-1} \tag{5}$$

$$K_k = \frac{P_{k|k-1}H_k^T}{H_k P_{k|k-1}H_k^T + R_k} \tag{6}$$

$$x_{k|k} = \hat{x}_{k|k-1} + K_k(z_k - H_k x_{k|k-1}) \tag{7}$$

$$P_{k|k} = P_{k|k-1} - K_k H_k P_{k|k-1} \tag{8}$$

Among them, formula (4) marks the state prediction equation, $\hat{x}_{k-1}$ corresponds to the optimal estimated state at $k-1$, and $\hat{x}_{k|k-1}$ is the predicted state based on the optimal estimation. Equation (5) stands for the prediction method of error covariance, P_{k-1} corresponds to the updated error covariance matrix at $k-1$, and $P_{k|k-1}$ refers to the predicted value of error covariance. Equation (6) marks the Kalman gain calculation method, and K_k corresponds to the Kalman gain value at time k that is used to subsequently optimize the target state prediction value and the error covariance prediction value. Equation (7) represents the target state optimization method, and $x_{k|k}$ corresponds to the optimal predicted value at time k. Equation (8) means the optimization method of error covariance, and $P_{k|k}$ corresponds to the optimal estimated value at time k.

In general, the movement of the same target among adjacent frames can be approximated as a uniform linear movement. Then the target width W, height H, speed V and interval time Δt have the following relationship:

$$W_k = W_{k-1} + \Delta t * V_{W|k-1} \tag{9}$$

$$H_k = H_{k-1} + \Delta t * V_{H|k-1} \tag{10}$$

$$V_{W|k} = V_{W|k-1} \tag{11}$$

$$V_{H|k} = V_{H|k-1} \tag{12}$$

The state vector x needs to record speed and size information, which can be expressed as:

$$x = (W_K, H_K, V_{W|k}, V_{H|k})^T \tag{13}$$

Among them, W_k and H_k are the width and height of the target, $V_{W|k}$ and $V_{H|k}$ mark the components of the target's speed in width and height. Combined with the above relational equation, Δt is the unit time, and the state transition matrix can be expressed as follows:

$$A = \begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{14}$$

Since the position and size in the state vector can be directly used as the value in the observation vector, the observation matrix is expressed as follows:

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \tag{15}$$

The corresponding values of the system noise covariance matrix Q and the observation noise covariance matrix R are as follows:

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{16}$$

$$R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{17}$$

In the initialization phase of the Kalman filter, the position and area of the target can be obtained during the target detection process. Initially setting the target's velocity change rate to 0. When considering the large randomness caused by initialization, the initial covariance matrix is assigned as follows:

$$P=\begin{pmatrix}100 & 0 & 0 & 0\\ 0 & 100 & 0 & 0\\ 0 & 0 & 100 & 0\\ 0 & 0 & 0 & 100\end{pmatrix} \quad (18)$$

After the Kalman filter is initialized, the state of the target can be predicted and optimally estimated in an iterative process based on the subsequent target detection information. This iterative process corresponds to the target tracking process.

3.3 Establishment of Monocular Vision Ranging Model

3.3.1 Ranging Principle

The central perspective projection imaging model is also called the pinhole imaging model. The model assumes that the reflected or emitted light from the surface of the object is projected onto the image plane through a "pinhole" that is called the optical center C and the projection center. As shown in Figure 1, the schematic diagram of the principle of the small hole imaging model is the imaging principle diagram of the central perspective projection model. The pinhole imaging model is mainly composed of optical center, imaging plane and optical axis. Among them, the image distance v from the optical center to the image plane is called the focal length f, and the object distance u is equal to the distance from the optical center to the object.

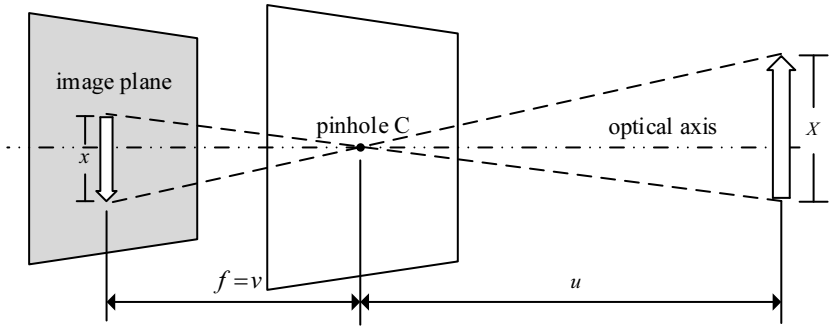


Figure 1. Principle of monocular ranging

According to the imaging process of the central perspective projection model, the distance X from the target to the optical axis and the distance x from the image plane to the optical axis satisfy the linear relationship of similar triangles in geometric optics, as displayed in formula (19):

$$\frac{X}{u}=\frac{x}{f} \quad (19)$$

3.3.2 Coordinate System Conversion

Projection in the three-dimensional world coordinates to the pixel value of the two-dimensional gray matrix with the cooperative symbol fixed by the geometric dimension. The most common phenomenon is the pinhole camera model. The conversion process is illustrated in Figure 2 below.

Both the world point and the camera are in a three-dimensional coordinate system, and only a proper rotation and translation are required to convert specific virtual world coordinates to camera-centric camera coordinates. The world coordinate of any point P is (X_w, Y_w, Z_w). The conversion from the world coordinate system to the camera coordinate system involves rotation and translation. The conversion is realized by the rotation of matrix R and the transformation of matrix T.

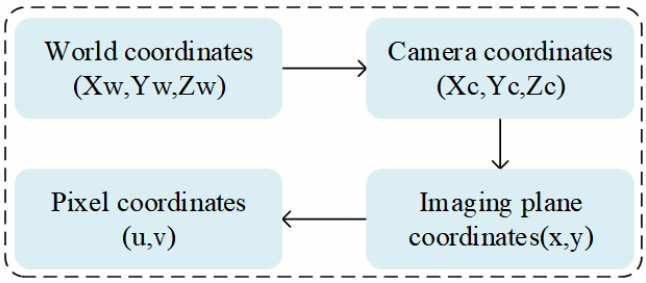


Figure 2. Coordinate system transformation flow chart

$$\begin{bmatrix}X_c\\ Y_c\\ Z_c\end{bmatrix}=R\begin{bmatrix}X_w\\ Y_w\\ Z_w\end{bmatrix}+T \quad (20)$$

From the camera coordinate system to the image coordinate system that belongs to the perspective projection relationship, when converting from 3D to 2D. As illustrated in formula (21).

$$Z_c\begin{bmatrix}x\\ y\\ 1\end{bmatrix}=\begin{bmatrix}f_x & 0 & 0 & 0\\ 0 & f_y & 0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix}\begin{bmatrix}X_c\\ Y_c\\ Z_c\\ 1\end{bmatrix} \quad (21)$$

Both pixel coordinate system and image coordinate

system are on the imaging plane, while their origin and measurement units are different. The origin of the image coordinate system is the intersection of the camera optical axis and the imaging plane, which is usually the principal point of the imaging plane. The unit of image coordinate system is millimeter, and belongs to physical unit, and the unit of pixel coordinate system is pixel. Therefore, the conversion between the two is demonstrated in formula (22), where *dx* and *dy* represent the number of millimeters each column and row represents. The pixel coordinate origin is (u_0, v_0), and the image coordinate of the arbitrary coordinates (u, v).

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & 0 \\ 0 & \frac{1}{dy} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{22}$$

In summary, the relationship between the conversion of the world coordinate system and the pixel coordinate system is displayed in formula (23).

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{23}$$

3.3.3 KFPA Monocular Ranging Method

When the distance is determined, the pixel area of the larger object in the picture is relatively large. For the same target, a longer distance refers to that the pixel of the object is smaller. Therefore, for a fixed target, the distance is inversely proportional to the pixel size of the object in the image. Based on the pinhole imaging model, it is proposed to introduce the pixel area as the basis of ranging, so as to improve the stability of ranging.

On the contrary, after the target detection step, the target detection result (x, y) is obtained, where x is the pixel width of target detection, and y represents the pixel height of target detection. X stands for the actual width of the target, and Y marks the actual height of the target. f_x indicates the focal length of the camera in the x direction, and f_y indicates the focal length of the camera in the y direction. Based on the pinhole imaging model:

$$x = \frac{X * f_x}{u} \tag{24}$$

$$y = \frac{Y * f_y}{u} \tag{25}$$

The relationship between pixel area and distance is as follows:

$$s = xy = \frac{f_x f_y XY}{u^2} \tag{26}$$

$$u = \sqrt{\frac{f_x f_y XY}{s}} \tag{27}$$

During the experiment, the focal length f often is a fixed value. The actual width and height (X,Y) of the target is also a fixed value using the mean value. After Kalman filter processes, the optimized target width and height (W_c, H_c) are output. Based on the relationship between pixel area and distance, this algorithm adds a ranging method on the basis of aspect ratio. Solve the influence of multiple angle changes and improve the applicability. The algorithm 2 is as follows:

Algorithm 2. Multi-angle Adaptive KFPA Monocular Ranging Method.

Require:

State vector $x = (W_k, H_k, v_w, v_H)^T$

Target actual width and height (X,Y)

Ratio of width to height R

Ensure:

Distance u

1. while the program does not stop do
2. $\hat{x}_{k|k-1} = A_{k|k-1}\hat{x}_{k-1}$
3. $P_{k|k-1} = A_{k|k-1}P_{k-1}A_{k|k-1}^T A_{k|k-1}^T + Q_{k-1}$
4. $K_k = P_{k|k-1}H_k^T / H_k P_{k|k-1} H_k^T + R_k$
5. $x_{k|k} = \hat{x}_{k|k-1} + K_k(z_k - H_k x_{k|k-1})$
6. $P_{k|k} = P_{k|k-1} - K_k H_k P_{k|k-1}$
7. $W = x_{k|k}.W_k$
8. $H = x_{k|k}.H_k$
9. $R_c = W/H$
10. if $R_c < R - a$ then
11. $H_c = H$
12. $W_c = H_c * R$
13. else if $R_c > R + a$ then
14. $W_c = W$
15. $H_c = W_C / R$
16. else
17. $W_c = W$
18. $H_c = H$
19. end if
20. $u = \sqrt{\frac{f_x f_y XY}{W_c H_c}}$
21. end while

Algorithm establishes the corresponding relationship between the area and distance of the target pixel. When the shooting angle changes, the pixel width and height of the front target are deduced by comparing the front aspect ratio R with the actual aspect ratio R_c. Kalman filter is introduced to deal with tracking fluctuation, which improves the stability of ranging.

4 Experimental Platform

The real-time target tracking mobile experimental platform is mainly composed of Cambrian 1H8, raspberry pi 3B, stm-32 and other hardware. The real object of the smart car is listed in Figure 3:

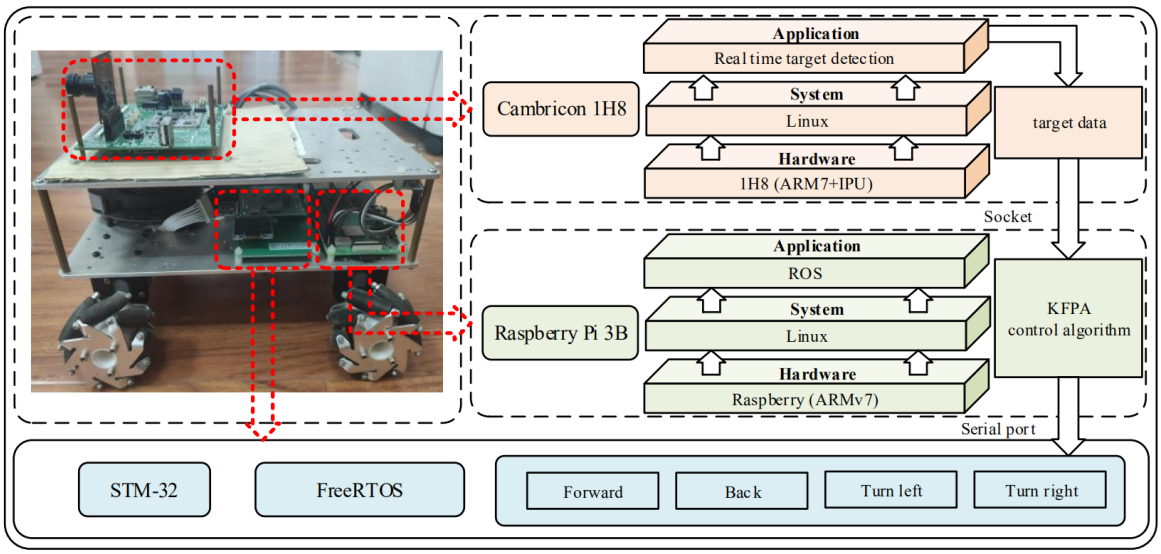


Figure 3. Target tracking smart car architecture

Cambrian 1H8 is regarded as a real-time video capture module. Its core is the integrated Cambrian 1H8, the processor's main frequency can reach 1GHz, and the power consumption is only 1.2W. Cambrian 1H8 can be widely applied in the field of computer vision to improve the visual grasping and real-time processing capabilities. In this experiment, the application running on 1H8 detects the real-time objects and transmits the data to the Raspberry Pi 3B.

Raspberry Pi 3B is utilized as an industrial computer for smart cars. It is configured with ARMv7, 1GB RAM, SD 16GB and Ubuntu mate system. Robot Operating System (ROS) is installed on Raspberry Pi that receives the target detection data of the upper layer 1H8 through the socket. The data is processed by KFPA algorithm and control algorithm, so as to generate speed data and transmit to STM-32.

STM-32 is applied to directly control the hardware. When STM-32 receives the angular velocity and linear velocity from the serial port, it adds corresponding voltage to the physical motor, thus realizing the movement of the smart car. And then, the smart car can track the target in real time.

5 Experiment

5.1 Parameter Calibration

In this verification experiment, first the distance D between the camera and the object with a fixed object is changed, and photos corresponding to different distances are taken. Subsequently, the data of pixel width and height corresponding to different distances can be obtained by manual labeling. Finally, these data are applied to fit the pixel area and distance curve, and the fitting result of matlab cftool is displayed in Figure 4. Goodness of Fit R^2=0.9999 .

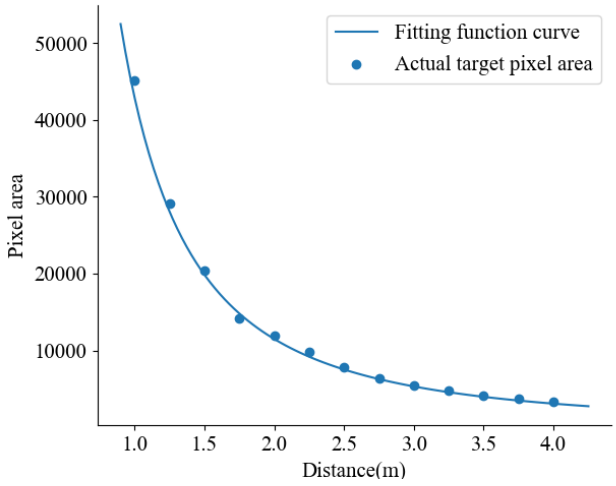


Figure 4. Camera parameter calibration

This is an inverse quadratic function and also verifies that the previous speculation is correct. The equation was tested with another set of data from the same camera, and a better fitting result was obtained. The parameter calibration test object is wide and high (0.38m, 0.48m), and the distance measurement formula based on pixel area is as follows:

$$s = 2.398 * 10^5 * x^{-2} \quad \textbf{(28)}$$

5.2 Static Monocular Ranging

5.2.1 Fixed Angle Change Distance

The optical axis of the camera is parallel to the ground and remain still. The actual distance between the experimental human body and the camera is changed. The experimental results of actual distance and calculated distance are exhibited in Table 1.

Table 1. Fixed angle change distance

Num	Pixel area	Actual distance (cm)	Calculate distance (cm)	Error (%)
1	32533.53	250	271.4	8.56
2	30854.77	275	278.7	1.35
3	26883.61	300	298.6	0.47
4	20284.81	325	340.8	4.62
5	17884.31	350	366.1	4.60
6	15963.73	375	387.5	3.33
7	15137.45	400	398.0	0.5
8	13006.46	425	429.3	1.01
9	11080.78	450	465.1	3.36
10	10291.51	475	482.7	1.62
11	9760.07	500	495.6	0.88
12	9189.83	525	510.8	2.70
13	7877.28	550	551.7	0.31
14	7159.28	575	578.7	0.64
15	6451.96	600	609.6	1.60
16	6157.56	625	624.0	0.16
17	5807.32	650	642.5	1.15
18	5232.53	675	676.9	0.28
19	4882.89	700	700.7	0.10

Table 2. Fixed distance change angle

Num	Angle	Pixel area only	Error (%)	Area and aspect ratio	Error (%)
1	10	268.4	3.23	267.9	3.04
2	20	268.0	3.08	265.4	2.08
3	30	266.8	2.62	260.5	0.19
4	40	273.1	5.04	258.1	0.73
5	50	288.7	11.04	267.9	3.04
6	60	306.2	17.77	270.4	4.00

Experimental results reveal that when the distance between the camera and the moving object is within the range of 2.5-7 meters, the measurement error is less than 5%. The larger error at 2.5 meter is due to inaccurate target detection. The reason is that the longer the distance, the smaller the target pixel area. When the distance exceeds 10, the error would increase significantly. This method is suitable for target detection and ranging is within 10 meters.

5.2.2 Fixed Distance Change Angle

Keeping the distance between the camera and the target at 2.6m. The distance calibration is the linear distance D from the camera to the target center. Changing the angle between the optical axis of the camera and the horizontal plane by 10°-60°. The experimental results are illustrated in Table 2.

This experiment is based on the frontal perspective of pedestrians, and changing the camera angle 0-60°. Due to the fluctuation of target detection, the improvement of the ranging effect based on area and aspect ratio is not obvious. However, when the detection angle is relatively large, the pixel area differs greatly from the horizontal viewing angle area, which greatly improves the ranging effect on the basis of area and aspect ratio. It can provide more effective ranging information for Smart car/drone with changing perspectives.

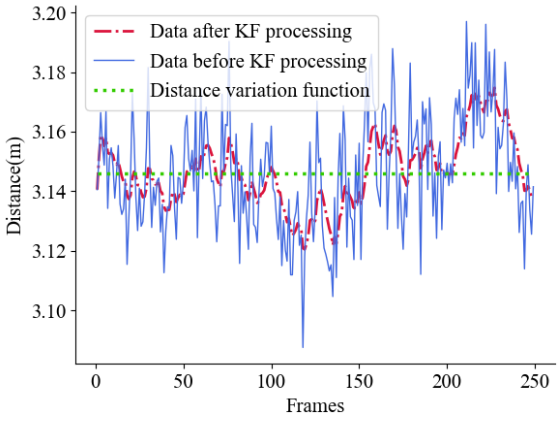
5.3 Dynamic Real-time Monocular Ranging

The ranging equipment is ultimately mounted on mobile platforms, including smart cars and drones, because the mobility of the detection target and the carrier platform could affect the accuracy of real-time target detection, and then affect the effect of ranging. Introducing Kalman filter can improve the stability of real-time ranging based on target detection. The dynamic real-time ranging experiment is divided into three groups: the distance between vehicle and target remains unchanged, the distance between vehicle and

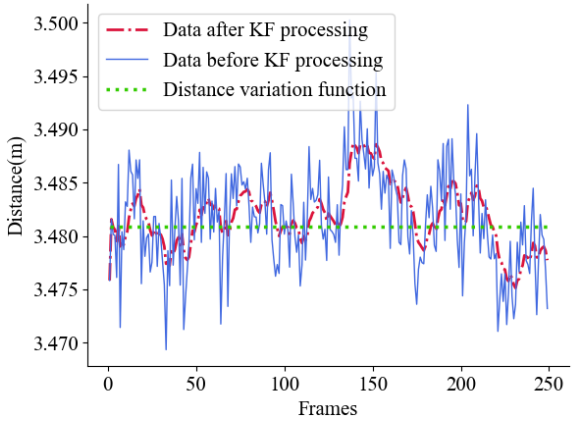
target becomes larger, and the distance between vehicle and target becomes smaller.

5.3.1 The Distance Between Vehicle and Target Remains Unchanged

Figure 5(a) and Figure 5(b) show the real-time ranging effect when the target and platform are relatively stationary. Subgraph Figure 5(a) displays that the target and the vehicle are moving forward, and the speed of the vehicle is equal to the speed of the target $v_{car} \approx v_{target}$. Subgraph Figure 5(b) exhibits that the target and the vehicle are moving backward, and the speed of the vehicle is equal to the speed of the target $v_{car} \approx v_{target}$.



(a) Forword $v_{car} \approx v_{target}$



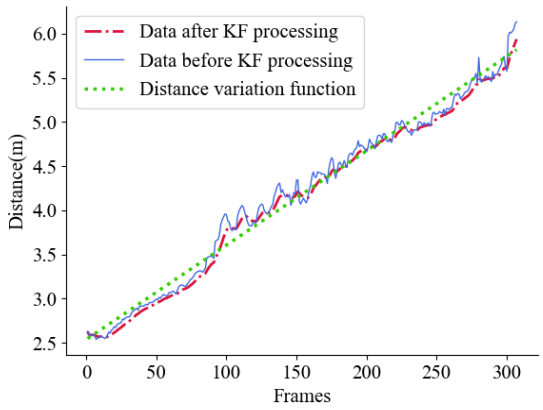
(b) Back off $v_{car} \approx v_{target}$

Figure 5. Relative static experiment

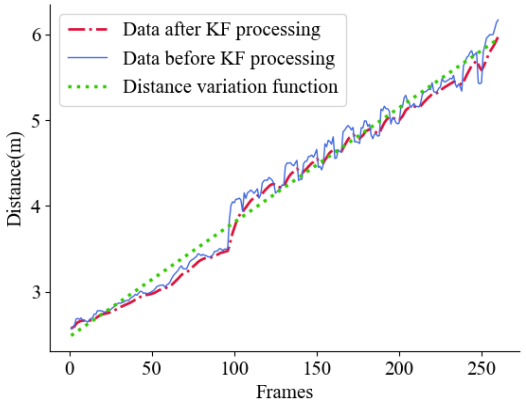
When the target and the carrier platform are relatively stationary at the same speed, real-time target detection causes some fluctuations in the ranging. Variance is adopted to evaluate the stability of real-time ranging. In Figure 5(a), the green line represents the average linear change trend of the number of frames and the distance, and is used to calculate the stability of the data before and after filtering. The blue line suggests that the variance of the unfiltered wave is 0.0191, and the red refers to that the variance of the filtered wave is 0.0116. The variance after filtering is reduced by 39.27% than before. After taking the average value of several groups of experiments, the stability increased by 25.47%.

5.3.2 The Distance Between Vehicle and Target Becomes Larger

Figure 6(a) and Figure 6(b) illustrates that the distance between the target and the platform is increasing over time. Subgraph Figure 6(a) demonstrates that the target and the vehicle are moving forward, and the vehicle speed is less than the target speed $v_{car} < v_{target}$. Subgraph Figure 6(b) gives information that the target and the vehicle are moving backward, and the vehicle speed is greater than the target speed $v_{car} > v_{target}$.



(a) Forword $v_{car} < v_{target}$



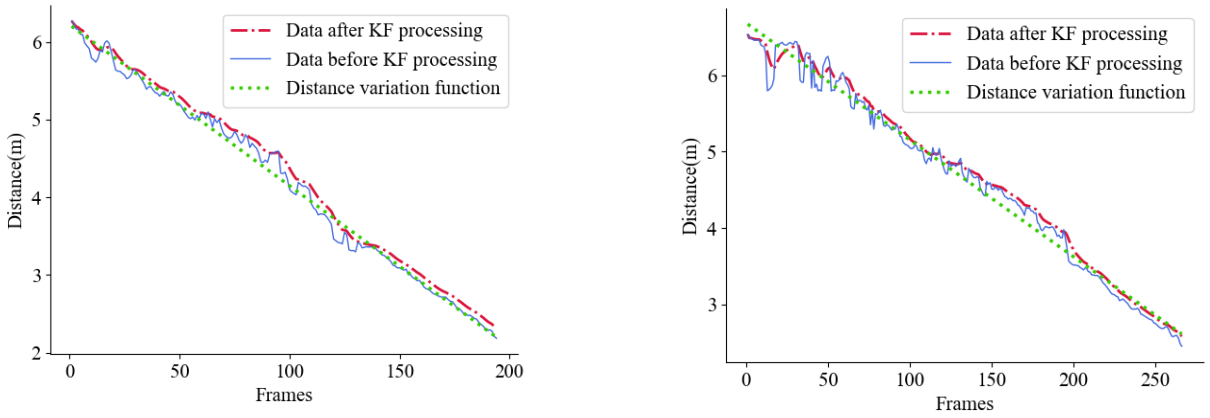
(b) Back off $v_{car} > v_{target}$

Figure 6. Distance increase experiment

When the distance between the target and the platform continues to increase over time. Variance is adopted to evaluate the stability of real-time ranging. In Figure 6(a), the blue line indicates that the unfiltered wave has a variance of 0.1440, and the red line means that the filtered wave has a variance of 0.1179. The variance after filtering is reduced by 18.12% compared with before. After taking the average value of several groups of experiments, the stability increased by 26.58%.

5.3.3 The Distance Between Vehicle and Target Becomes Smaller

Figure 7 shows that the distance between the target and the platform is decreasing over time. Subgraph Figure 7(a) exhibits that the target and the vehicle are moving forward, and the vehicle speed is greater than the target speed $v_{car} > v_{target}$. Subgraph Figure 7(b) concludes that the target and the vehicle are moving backward, and the vehicle speed is less than the target speed $v_{car} < v_{target}$.



(a) Forword $v_{car} > v_{target}$ (b) Back off $v_{car} < v_{target}$

Figure 7. Distance reduction experiment

When the distance between the target and the platform continues to decrease over time. Variance is applied to evaluate the stability of real-time ranging. In Figure 7(a), the blue line means that the unfiltered wave has a variance of 0.1096, and the red line marks that the filtered wave has a variance of 0.0861. The variance after filtering is reduced by 21.59% than before. After taking the average value of several groups of experiments, the stability increased by 23.58%.

6 Conclusion

In this paper, we studied the monocular vision ranging method in the image and proposed a monocular ranging method based on the target pixel area and aspect ratio, so as to accurately obtain the distance between the camera and the target. First, we utilize the pixel area as one of the elements of distance measurement, which effectively improves the stability compared with the method that only depends on width or height. Secondly, this method also adds the target aspect ratio, thus improving the detection accuracy of the camera and target view angle changes. Finally, in the process of target detection data, this method introduces Kalman filter process to improve the stability of real-time detection as a whole. In the pedestrian target detection experiment, the experimental results prove that, when the distance is within 3-7m, the average relative error of the distance measurement is less than 4.62%. On the other hand, when the angle transformation is 10°-60°, the average relative error of distance measurement is less than 4%.

This ranging method combines KCF-SIFT target detection and tracking method together. Real-time tracking of simple targets is achieved on the ROS smart car platform. This method has made effective contributions to the construction of a multivehicle collaborative platform.

Acknowledgements

The work is supported by the National Key Research and Development Program under Grant No. 2019YFC0118404, the National Natural Science Foundation of China under Grant No. U20A20386, the Zhejiang Key Research and Development Program under Grant No. 2020C01050, the Key Laboratory fund general project under Grant No. 6142110190406, the Zhejiang Natural Science Foundation Project under Grant No. LY19F020044.

References

[1] J. Straub, Unmanned aerial systems: Consideration of the use

of force for law enforcement applications, *Technology in Society*, Vol. 39, No. 1, pp. 100-109, November, 2014.

[2] H. H. Gao, C. Liu, Y. H. Z. Li, X. X. Yang, V2VR: Reliable Hybrid-Network-Oriented V2V Data Transmission and Routing Considering RSUs and Connectivity Probability, *IEEE Transactions on Intelligent Transportation Systems*, pp. 1-14, April, 2020, DOI: 10.1109/TITS.2020.2983835.

[3] Y. Y. Yin, Z. X. Cao, Y. S. Xu, H. H. Gao, R. Li, Z. Mai, QoS Prediction for Service Recommendation With Features Learning in Mobile Edge Computing Environment, *IEEE Transactions on Cognitive Communications and Networking*, Vol. 6, No. 4, pp. 1136-1145, December, 2020.

[4] G. Y. Jia, G. J. Han, A. H. Li, J. X. Du, SSL: Smart Street Lamp Based on Fog Computing for Smarter Cities, *IEEE Transactions on Industrial Informatics*, Vol. 14, No. 11, pp. 4995-5004, November, 2018.

[5] G. Y. Jia, G. J. Han, H. L. Rao, L. Shu, Edge Computing-based Intelligent Manhole Cover Management System for Smart Cities, *IEEE Internet of Things Journal*, Vol. 5, No. 3, pp. 1648-1656, June, 2018.

[6] S. G. Deng, Z. Z. Xiang, P. Zhao, J. Taheri, H. H. Gao, J. W. Yin, A. Y. Zomaya, Dynamical Resource Allocation in Edge for Trustable Internet-of-Things Systems: a Reinforcement Learning Method, *IEEE Transactions on Industrial Informatics*, Vol. 16, No. 9, pp. 6103-6113, September, 2020.

[7] M. O. Khyam, S. S. Ge, X. Li, M. R. Pickering, Highly accurate time-of-flight measurement technique based on phase-correlation for ultrasonic ranging, *IEEE sensors journal*, Vol. 17, No. 2, pp. 434-443, January, 2017.

[8] F. Rodriguez-Morales, C. Leuschen, C. L. Carabajal, J. Paden, J. A. Wolf, S. Garrison, J. W. McDaniel, An improved UWB microwave radar for very long-range measurements of snow cover, *IEEE Transactions on Instrumentation and Measurement*, Vol. 69, No. 10, pp. 7761-7772, October, 2020.

[9] L. P. Li, Z. Su, L. Xie, X. Zhao, A Laser Radar with High Precision Ranging based on Phase, *3rd Workshop on Advanced Research and Technology in Industry (WARTIA 2017)*, Guilin, China, 2017, pp.434-447.

[10] Y. Kang, Z. Cui, Design of Tram Collision Prevention System Based on Infrared Ranging and Kalman Filtering, *Journal of Physics: Conference Series*, Vol. 1237, No. 3, Article No. 032084, June, 2019.

[11] G. Y. Jia, G. J. Han, J. F. Jiang, L. Liu, L. Shu, DPAM: A Demand-based Page-level Address Mappings Algorithm in Flash Memory for Smart Industrial Edge Devices, *IEEE Transactions on Industrial Informatics*, Vol. 16, No. 3, pp. 1993-2002, March, 2020.

[12] H. H. Gao, W. Q. Huang, Y. C. Duan, The Cloud-edge-based Dynamic Reconfiguration to Service Workflow for Mobile Ecommerce Environments: A QoS Prediction Perspective, *ACM Transactions on Internet Technology (TOIT)*, Vol. 21, No. 1, pp. 1-23, February, 2021.

[13] G. Y. Jia, G. J. Han, J. X. Du, S. Chan, PMS: Intelligent Pollution Monitoring System Based on the Industrial Internet of Things for a Healthier City, *IEEE Network*, Vol. 33, No. 5, pp. 34-40, September-October, 2019.

[14] Y. Yu, X. X. Kang, J. H. Song, Pedestrian Recognition and Rapid Ranging Method of Night Based on Monocular Vision, *2019 Chinese Control And Decision Conference (CCDC)*, Nanchang, China, 2019, pp.5627-5632.

[15] S. X. Guo, S. Z. Chen, F. G. Liu, X. F. Ye, H. B. Yang, Binocular vision-based underwater ranging methods, *2017 IEEE International Conference on Mechatronics and Automation (ICMA)*, Takamatsu, Japan, 2017, pp. 1058-1063.

[16] Y. Qian, Q. Chen, W. G. Yang, F. Wang, W. X. Qian, Z. X. Li, Monocular ranging system based on space geometry, *Optics and Photonics for Information Processing XIII*, Vol. 11136, Article No. 111360Y, September, 2019.

[17] H. H. Gao, L. Kuang, Y. Y. Yin, B. Guo, K. Dou, Mining consuming Behaviors with Temporal Evolution for Personalized Recommendation in Mobile Marketing Apps, *Mobile Networks and Applications*, Vol. 25, No. 4, pp. 1233-1248, August, 2020.

[18] X. X. Yang, S. J. Zhou, M. Cao, An Approach to Alleviate the Sparsity Problem of Hybrid Collaborative Filtering Based Recommendations: The Product-Attribute Perspective from User Reviews, *Mobile Networks and Applications*, Vol. 25, No. 2, pp. 376-390, April, 2020.

[19] Q. L. Wang, Y. Zhang, W. C. Shi, and M. Nie, Laser ranging-assisted binocular visual sensor tracking system, *Sensors*, Vol. 20, No. 3, Article No. 688, February, 2020.

[20] M. J. Xiong, H. M. Lu, D. Xiong, J. H. Xiao, M. Lv, Scale-aware monocular visual-inertial pose estimation for aerial robots, *2017 Chinese Automation Congress (CAC)*, Jinan, China, 2017, pp. 7030-7034.

[21] S. H. Meng, A. C. Huang, T. J. Huang, Z. M. Cai, Q. Z. Ye, F. M. Zou, Ultrasonic ranging-based vehicle collision avoidance system, *Proceeding of the Twelfth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Kaohsiung, Taiwan, 2016, pp. 211-218.

[22] Z. Yang, C. Xu, B. Q. Deng, S. Q. Zeng, S. H. Yang, Q. P. Li, Z. B. Yang, Progress in the Application of Ranging Technology in Intelligent Driving, *3rd International Conference on Information Technology and Industrial Automation*, Guangzhou, China, 2018, pp. 204-211.

[23] B. Li, X. L. Zhang, M. Sato, Pitch angle estimation using a Vehicle-Mounted monocular camera for range measurement, *2014 12th International Conference on Signal Processing (ICSP)*, Hangzhou, China, 2014, pp. 1161-1168.

[24] M. Rezaei, M. Terauchi, R. Klette, Robust vehicle detection and distance estimation under challenging lighting conditions, *IEEE transactions on intelligent transportation systems*, Vol. 16, No. 5, pp. 2723-2743, October, 2015.

[25] R. Adamshuk, D. Carvalho, J. H. Z. Neme, E. Margraf, S. Okida, A. Tusset, M. M. Santos, R. Amaral, A. Ventura, S. Carvalho, On the applicability of inverse perspective mapping for the forward distance estimation based on the HSV colormap, *2017 IEEE International Conference on Industrial Technology (ICIT)*, Toronto, ON, Canada, 2017, pp. 1036-1041.

[26] H. W. Ho, G. C. de Croon, Q. P. Chu, Distance and velocity estimation using optical flow from a monocular camera,

International Journal of Micro Air Vehicles, Vol. 9, No. 3, pp. 198-208, September, 2017.

[27] C. Xu, M. Gao, and H. Cao, The Attitude Angle Estimation-based Distance Measurement of Tank Target in Monocular Image, *Acta Photonica Sinica*, Vol. 44, No. 5, Article No. 0512002, May, 2015.

[28] J. Han, O. Heo, M. Park, S. Kee, and M. Sunwoo, Vehicle distance estimation using a mono-camera for FCW/AEB systems, *International journal of automotive technology*, Vol. 17, No. 3, pp. 483-491, June, 2016.

[29] D. S. Bao, P. K. Wang, Vehicle distance detection based on monocular vision, *2016 International Conference on Progress in Informatics and Computing (PIC)*, Shanghai, China, 2016, pp. 187-191.

[30] M. M. Wang, Y. Liu, Z. Y. Huang, Large margin object tracking with circulant feature maps, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, Hawaii, USA, 2017, pp. 4800-4808.

Biographies

Shuo Chen is currently pursuing the master's degree with the Department of Computer Science, Hangzhou Dianzi University, China. His current research interests include the mobile edge computing and intelligent system.

Songzhu Mei is an assistant researcher at the National Defense Science and Technology Key Laboratory of Parallel and Distributed Processing, School of Computer Science, National University of Defense Technology. His main research direction is big data analysis and performance optimization.

Gangyong Jia received the Ph.D. degree from the Department of Computer Science, University of Science and Technology of China, Hefei, China, in 2013. He is currently an Associate Professor with the Department of Computer Science, Hangzhou Dianzi University, China. His current research interests include the IoT, cloud computing, edge computing, power management, and operating system.

Youhuizi Li received the Ph.D. degree in computer science from Wayne State University, Detroit, MI, USA, in 2016. She is currently an Assistant Professor with the Department of Computer Science, Hangzhou Dianzi University, China. Her current research interests include energy efficient computing, mobile and Internet computing, and big data systems.

Weihua Zhao is currently an Associate Professor with the Department of Computer Science, Hangzhou Dianzi University, China. Her current research interests include the edge computing, operating system, and power management.